

"Автоматическая переработка текста методами прикладной лингвистики". Чилинцев, 1971.

3. Морозенко В.В., Турук И.Ф. Лексический минимум по английскому языку для технических вузов. М., 1975.
4. Смирнова Л.С. Методические основы оптимизации обучения чтению литературы по специальности с учетом статистических характеристик текста (английский язык). Канд. дис. М., 1977.

В.К. КОЛОБАЕВ
Ленинградский санитарно-
гигиенический медицинский
институт

ОПЫТ СОСТАВЛЕНИЯ ЧАСТОТНОГО СЛОВАРЯ

Оптимизация процесса обучения иностранным языкам в вузах неязыкового профиля требует строгого отбора материала, как лексического, так и грамматического, в соответствии с целями и задачами, определенными программой обучения. В связи с этим становится ясной та важная роль, которую может сыграть частотный словарь в создании учебников и учебных пособий для таких вузов. В настоящее время проводится большая работа по составлению частотных словарей разного типа с различными установками (см. напр. [1]). Однако следует отметить, что остается еще немало областей науки, язык которых не подвергался статистической обработке. Это создает определенные трудности при обучении. К числу таких областей относится, в частности, гигиена.

Помимо чисто прикладной цели, создание частотных словарей позволяет определить некоторые лексические и синтаксические особенности исследуемого подязыка науки, выявить наиболее характерные для него слова, обороты, грамматические и морфологические формы. Исходя из данных частотного словаря, можно, по необходимости, выделить различные слои лексики, в частности, можно определить соотношения терминологической и общепотребительной лексики, количество заимствований и слов латинско-греческого происхождения. Последний аспект представляется особенно важным при работе в медицинском вузе, поскольку большое коли-

чество терминов (какое именно — должен показать создаваемый словарь) являются именно латинского или греческого происхождения. Все это делает работу по составлению частотных словарей не только актуальной, но и насущной потребностью при работе в языковом вузе.

В настоящее время в Ленинградском санитарно-гигиеническом медицинском институте ведется работа по созданию первого частотного англо-русского словаря гигиенического профиля. За основу создания словаря принята методика, разработанная П.М.Алексеевым и Н.Б.Карачан [2]. Для работы над словарем на кафедре иностранных языков создана группа из семи преподавателей. Перспективный пятилетний план работы подразделяется на две части. Первая рассчитана на три года и включает: 1) создание модели специальности, 2) отбор литературы к исследованию, 3) расписывание текстов на карточки и создание картотеки. Во вторую часть входят статистическая обработка материала, составление словарной статьи и частотного списка словаря-минимума по обследуемой области науки (гигиене).

Модель специальности. Каждая область научной деятельности распадается на ряд подобластей, или дисциплин, поэтому важно установить их процентное соотношение с целью определения объема материала, который необходимо проанализировать в каждой из них, для их объективного отражения в словаре. Для этого на первом этапе работы над словарем была составлена модель специальности. Под моделью специальности понимается в данном случае набор всех основных дисциплин, которые входят в общую область гигиены и которые относятся к ней как частное к общему.

Модель специальности была построена на основе статистического анализа смежнейшей литературы по данному научному профилю. Вся область гигиены принимается за 100%. Затем определяется, какой процент в этом объеме составляет каждая из основных гигиенических дисциплин. Полученные результаты приводятся на рис. 1.

Как уже указывалось в литературе (см. работы П.М.Алексеева), для составления достаточно точного и полного частотного словаря необходимо проанализировать материал в 200 000 словопотреблений, или 200 текстов по 1000 слов каждый. В соответствии с полученной моделью специальности было проведено определение объема материала по каждой подобласти, который следовало



Рис. 1. Модель специальности гигиениста

проанализировать и расписать на карточки. Составление модели специальности сокращает объем работы по словарю, не влияя при этом на его точность и репрезентативность.

При отборе материала к исследованию были введены следующие требования: 1) отбор научной литературы только по узкой специальности, опубликованной за последние 15 лет; 2) точное определение объема последуемого материала по каждой подобласти; 3) в каждой подобласти использование текстов не менее чем трех авторов; 4) разбивка текстов на подтексты по 1000 словоупотреблений каждый.

Процедура разбивки текста на отрезки по 1000 словоупотребления выглядит следующим образом. Для анализа берется, скажем, каждая пятая страница текста. На каждой из взятых для анализа страниц подсчитывается количество слов в самой верхней и самой

нижней полных строках. Полученные числа выписываются отдельно. Затем учитывается среднее количество слов в строке. Это среднее число будет содержать десятые и сотые доли, которые не отбрасываются и не округляются до целых чисел. Затем определяется количество строк, в которых содержится 1000 словоупотреблений (1000 делится на среднее количество слов в строке). После этого отсчитывается количество строк, содержащих 1000 словоупотреблений. Это будет первый текст, подлежащий расписыванию на карточки.

В ходе работы над частотным словарем гигиенического профиля выписываются только существительные, прилагательные, глаголы и наречия. Предполагается, что остальные лексические единицы, которые относятся к разряду вспомогательной лексики, обладают такой же частотностью, как и в других областях науки, которые уже были подвергнуты статистической обработке, и данные относительно их частотности могут быть позаимствованы из имеющихся словарей подобного типа.

При работе над текстом лексический материал сначала расписывается на сводную карту. Каждое новое слово выписывается в графу, соответствующую начальной букве, и против него ставится цифра I, обозначающая, что слово встретилось один раз. Если слово встречается в тексте еще раз, то ставится еще одна цифра I и т.д. Для каждого текста, таким образом, составляется алфавитно-частотный список всех учитываемых единиц. Затем полученные результаты переносятся на карточки с указанием частотности слова в тексте и его контекстуального русского эквивалента.

Карточка строится следующим образом (см. рис. 2): в верхней части выписывается слово, рядом указывается часть речи, затем — количество употреблений слова в расписываемом тексте и контекстуальный перевод. Если слово имеет несколько контекстуальных значений, то выписываются все, и указывается, сколько раз какой вариант встретился в тексте. Ниже приводится шифр, указывающий номер источника и номер выборки.

| | |
|-----------|---------------|
| strain, n | вид IIIIIIIII |
| 111111111 | |
| 9 | 9 |
| 1-3 | |

Рис. 2. Образец оформления карточки

Организация работы по такому принципу ведет к сокращению количества карточек и общего объема картотеки.

На заключительном этапе происходит статистическая обработка полученного материала: составляется сводная карта всех выписанных слов, в которой указывается, сколько раз слово встретилось в текстах и в каком количестве текстов, в каких подобластях, указывается его контекстуальный перевод и частота того или иного варианта, если таковых оказывается несколько.

Описанная в данной статье методика не претендует на универсальность. Однако ознакомление с ней может, по нашему мнению, помочь при работе над словарями подобного типа.

Литература

1. Есаулкова М.Т. Из опыта составления отраслевого частотного словаря. — В кн.: Вопросы лингвистики и методики преподавания иностранных языков. Свердловск, 1970.
2. Алексеев П.М. Статистическая лексикография. Л., 1975; Карача, Н.Б., Вашук Н.Я., Маркова Н.Л. и др. Из опыта составления частотных терминологических словарей. — В кн.: Иностранные языки на неспециальных факультетах. Л., 1978.